·论著·

基于人工智能算法的脑卒中溶栓药物精准治疗：真实世界研究

沈惠文[1]，林永忠[2]，陈淑良[3]，张立红[4]，马春野[2]，马得原[1]，张策[1*]

【摘要】 **背景** 缺血性脑卒中（IS）起病急，治疗时间窗窄，治疗效果的影响因素复杂，患者自身情况各异，因此治疗方式、给药种类、给药剂量、给药方式均会影响患者的溶栓效果。既往研究常利用统计方法分析溶栓效果的影响因素，人工智能算法在该方面的临床应用尚少见。**目的** 基于真实世界的数据，建立IS患者从患者一般特征、药物治疗方式到恢复效果的人工智能算法模型，实现患者的个体化溶栓药物精准治疗，为临床用药决策提供数据支持。**方法** 采用回顾性研究方式，从大连医科大学附属第二医院医渡云科研大数据服务器系统提取本院确诊为IS的患者（n=55 621）的临床信息，时间为2001-01-01至2021-12-31。依据纳入标准共筛选出信息完整的IS患者1 855例，依据每位患者入院与出院时美国国立卫生研究院卒中量表（NIHSS）评分差值评价患者溶栓效果，并将患者分为神经功能改善组（差值≥ 4分为神经功能改善，n=1 236）和对照组（差值<4分为为神经功能未改善，n=619）。经3位神经内科高级职称专家背对背推荐，并结合查阅的IS诊治指南及文献，整理可能与IS发作后溶栓效果相关的影响因素，归类为患者一般特征、用药指标、检查指标、检验指标、治疗方式5类。首先进行影响因素的单因素筛选，再利用主成分分析法对影响因素做降维处理。构建Logistic回归模型、支持向量机（SVM）、C5.0决策树、深度神经网络（DNN）及Wide&Deep模型，进行模型对比评价，比较不同模型对IS患者溶栓效果的预测情况，确定最佳模型，进而寻找模型的最优参数。将1 855例患者的临床信息进行分割处理，随机数为7和11，随机分为训练集（1 113例）、验证集（371例）、测试集（371例），其中训练集用来构建和训练模型以发现规律，验证集用来调整模型参数，测试集用来评价最终模型的泛化能力。应用特征工程构建简化模型并评估模型准确度。从大连市中心医院的医渡云科研大数据服务器系统中提取IS患者的临床信息（共提取3 925例），利用其数据进行外部验证。**结果** 共纳入26个患者特征（即溶栓效果影响因素）进行模型构建。经主成分分析降维成2个主成分，累计方差贡献率为93.1%。比较Logistic回归模型、SVM、C5.0决策树、DNN及Wide&Deep模型预测溶栓效果的价值，发现Wide&Deep模型的预测性能最佳，准确度为81.5%，F指数为87.1%。训练集ROC曲线下面积为0.753，测试集ROC曲线下面积为0.793。确定Wide&Deep模型的隐含层层数为7层，每层神经元为个数为15个，以Sigmoid作为激活函数，模型参数最优。IS患者溶栓治疗后神经功能改善影响因素的特征工程分析结果显示，用药种类、给药方式和用药剂量的重要性排序均在前列，重要性排序由大到小分别为：脑血管病史、用药种类、给药方式、单次剂量、动脉粥样硬化、溶栓时间窗、是否使用抗凝药物和活血化瘀药物等。模型自变量简化后得出，Wide&Deep模型准确度为0.819，模型自变量简化后外部验证的准确度为0.801。**结论** Wide&Deep模型各项评价指标优异，影响溶栓效果的因素排序由大到小分别为：脑血管病史、用药种类、给药方式、单次剂量、动脉粥样硬化、溶栓时间窗、是否使用抗凝药物和活血化瘀药物等。通过人工智能算法从影响因素和个体化给药方面可为临床医生提供及时和有效的IS患者药物溶栓治疗方案，对减轻疾病社会负担具有积极意义。

【关键词】 缺血性卒中；溶栓药物；人工智能算法；Wide&Deep模型；精准治疗

【中图分类号】 R 743.3 【文献标识码】 A DOI:10.12114/j.issn.1007-9572.2023.0048

沈惠文，林永忠，陈淑良，等. 基于人工智能算法的脑卒中溶栓药物精准治疗：真实世界研究［J］. 中国全科医学，2023.［Epub ahead of print］.［www.chinagp.net］

SHEN H W，LIN Y Z，CHEN S L，et al. Precise thrombolytic treatment for stroke using ai-based algorithms：a real-world study［J］. Chinese General Practice，2023.［Epub ahead of print］.

Precise Thrombolytic Treatment for Stroke Using AI-based Algorithms：a Real-world Study *SHEN Huiwen*[1]，*LIN*

基金项目：“1+X”计划临床研究孵化项目（2022LCYJYB05）

临床试验注册：ChiCTR2000040367

1.116027 辽宁省大连市，大连医科大附属第二医院发展规划与质量管理部 2.116027 辽宁省大连市，大连医科大学附属第二医院神经内科 3.116027 辽宁省大连市，大连医科大学附属第二医院护理部 4.116027 辽宁省大连市，大连市中心医院介入及神经重症科

*通信作者：张策，副主任药师；E-mail：zhangce@dmu.edu.cn

本文数字出版日期：2023-03-09

Yongzhong[2], *CHEN Shuliang*[3], *ZHANG Lihong*[4], *MA Chunye*[2], *MA Deyuan*[1], *ZHANG Ce*[1*]

1.Development Planning and Quality Management Department, the Second Affiliated Hospital of Dalian Medical University, Dalian 116027, China

2.Department of Neurology, the Second Affiliated Hospital of Dalian Medical University, Dalian 116027, China

3.Nursing Department, the Second Affiliated Hospital of Dalian Medical University, Dalian 116027, China

4.Department of Intervention and Neurocritical Care, Dalian Municipal Central Hospital, Dalian 116027, China

[*]Corresponding author: ZHANG Ce, Associate chief pharmaceutist; E-mail: zhangce@dmu.edu.cn

【Abstract】 **Background** The thrombolytic effect for ischemic stroke (IS) is affected by complex factors, such as acute onset of stroke, short therapeutic time window, various individual patient factors, treatment model, types and doses of medicines as well as mode of administration. To identify the influencing factors of thrombolytic effect, most existing studies adopt statistical methods, while rare studies use artificial intelligence (AI) -based algorithms. **Objective** To establish models using AI-based algorithms for IS patientsbased on the real-world data including general patient characteristics, medication model and recovery effects, to achieve precise individualized thrombolytic treatment and provide data support for clinical prescription decisions. **Methods** A retrospective design was used. The clinical information of IS patients (*n*=55 621) was extracted from the Yidu Cloud scientific research big data server system of the Second Affiliated Hospital of Dalian Medical University from January 1, 2001 to December 31, 2021, among whom1 855 with complete information were enrolled according to the inclusion criteria. Thrombolysis effect was evaluated by comparing the National Institutes of Health Stroke Scale (NIHSS) score measured at admission and discharge, and those with an improvement in the NIHSS score by ≥ 4 points and < 4 points were assigned to neurological improvement group (*n*=1 236), and control group (*n*=619), respectively. Factors possibly associated with post-IS thrombolytic effect (including general patient characteristics, medication indicators, examination indicators, test indicators, and treatment methods) were obtained by summarizing the factors suggested separately by three neurology experts with a senior title, and reviewing relevant guidelines and literature, then were screened using univariate analysis, and the identified ones were treated by dimensionality reduction using principal component analysis (PCA). Models of Logistic, support vector machine, C5.0 decision tree arithmetic, deep neural network, and Wide&Deep, were built and compared to find the one with the best performance in predicting thrombolytic effect, then to determine its parameters. Then by use of two randomly generated two numbers, 7 and 11, the 1 855 patients were randomly assigned to three datasets, training (*n*=1 113, for building and practicing models to discover rules), validation (*n*=371, for adjusting model parameters), and test (*n*=371, for evaluating the generalization ability of the final model). Feature engineering was used to construct a simplified model and evaluate its accuracy. The clinical information of IS patients (*n*=3 925) was extracted from theYidu Cloud scientific research big data server system of Dalian Central Hospital for external verification of the model. **Results** Twenty-six patients characteristics associated with thrombolytic effect were included for establishing models. The dimensionalities were reduced to two principal components by PCA, explaining 93.1% of the total variance. Comparison analysis revealed that the Wide&Deep model had the best predictive performance with an accuracy of 81.5%, and an F-index of 87.1%. Furthermore, the values of the area under the receiver operating characteristic (AUC) curve of the Wide&Deep model in predicting the thrombolytic effect in patients in the training set and test set were 0.753 and 0.793, respectively. The number of hidden layers and neurons in each layer of the model was seven and 15, respectively. Using sigmoid as the activation function showed that the model parameters were optimal. The feature-engineering analysis of factors influencing the improvement of neurological function showed that the importance of medication type, administration mode and dosage ranked high, and the importance ranking in a descending order is: cerebrovascular disease history, type of medication, mode of administration, single dose, atherosclerosis, therapeutic time window of thrombolytic therapy, prevalence of use of anticoagulant drugs and drugs for promoting blood circulation and removing blood stasis. After simplifying the independent variables of the model, the accuracy of the Wide&Deep model was 0.819, and its accuracy was 0.801 suggested by the external verification after model simplification, indicating good predictive performance and generalizability. **Conclusion** The Wide&Deep model has proven to have excellent evaluation indicators. The importance of influencing factors of thrombolyticeffect in a descending order is: cerebrovascular disease history, type of medication, administration mode, single dose, atherosclerosis, therapeutic time window of thrombolytic therapy, prevalence of use of anticoagulants and blood-activating and stasis-removing drugs. It provides clinicians with timely and effective thrombolysis treatment support involving thrombolysis related factors and individualized administration using AI-based algorithms.

【Key words】 Ischemic stroke; Thrombolytic drugs; Artificial intelligence algorithm; Wide&Deep model; Precision treatment

缺血性脑卒中（ischemic stroke，IS）起病急，治疗时间窗窄，治疗效果影响因素复杂，因此快速诊断患者病情、评估给予治疗方案是关键。现认为静脉溶栓是最主要的恢复 IS 患者血管血流循环的方式，常使用的静脉溶栓药物主要有阿替普酶和尿激酶，而患者自身情况各异，且治疗方式、给药种类、给药剂量、给药方式均会影响患者的溶栓效果。Wide&Deep 模型的核心是将深度神经网络具备的泛化性和线性模型的记忆性能相结合，目的是为了能够较大程度提升模型的预测性能[1]，同时解决“记忆”和“泛化”的问题。“记忆”是指在面对大量离散稀疏的特征时，将特征进行非线性转换，使其具有可解释程度强等优点。但随模型复杂程度增加，记忆能力增强，模型易出现过拟合的情况。而“泛化”恰好能缓解该类问题，“泛化”是稀疏特征学习低维的密集嵌入来捕获特征之间的相关性，即将高维向量转换到低维度空间中，需要较少的工作量。因此采用大数据的学习方式，根据已经采用某一方案溶栓的患者预后，反推干预手段，从而建立患者特征、干预手段包括混杂因素作为输入，预测治疗效果的大数据算法，包括 Wide&Deep、支持向量机（SVM）、分类回归树、C5.0 决策树、深度神经网络（DNN）等。

目前，医学大数据多集中在医学影像方面[2-4]，在临床辅助诊断和监测方面的研究尚少见。溶栓药物研究多集中于溶栓药物的临床效果观察[5-6]、用药效果及安全性分析[7-8]、溶栓效果影响因素[9-12]方面，样本量多为数十到数百不等。针对溶栓治疗方法的人工智能算法尚未见报道。本研究尽可能多的纳入混杂因素，如高血压、冠心病、用药史等既往病史，同时利用人工智能的有效算法，更全面的探讨了溶栓影响因素，也降低了混杂偏倚，以期在真实世界背景下，增加人工智能用药推荐模型的适用性。在 IS 患者在入院时，根据患者的不同特点，由模型算法，推荐最优治疗方法及给药方案，如用药剂量、用药种类及给药方式等，同时算法转化后应用于临床，以期对患者的识别、治疗方法的动态监测与调控、诊断操作的自动与精准提供有益探索。

1　对象与方法

1.1　研究对象　采用回顾性研究方式，从大连医科大学附属第二医院医渡云科研大数据服务器系统提取本院确诊为 IS 的患者（n=55 621）的临床信息，时间为 2001-01-01 至 2021-12-31。IS 患者纳入标准：（1）年龄 >18 岁，男女不限；（2）根据神经影像学诊断为有脑部存在责任缺血病灶；（3）患者入院及出院时美国国立卫生研究院卒中量表（NIHSS）评分完整；（4）具有溶栓指征。排除标准：（1）关键数据严重缺失，如用药剂量、治疗方式、NIHSS 评分等数据不能回溯；（2）伦理批件中规定为未脱敏对象。依据纳入标准筛选后，最终共纳入 IS 患者 1 855 例。

依据每位患者入院与出院时 NIHSS 评分差值评价患者溶栓效果，并将患者分为神经功能改善组（差值 ≥ 4 分为神经功能改善[13]，n=1 236）和对照组（差值 <4 分为为神经功能未改善，n=619）。

本研究经大连医科大学附属第二医院伦理委员会批准（大医二院伦字〔2020〕043）。

1.2　研究指标　经 3 位神经内科高级职称医师背对背推荐，并结合查阅 IS 诊治指南及文献，整理可能与 IS 发作后溶栓效果相关的影响因素，归类为患者一般特征、用药指标、检查指标、检验指标、治疗方式 5 类，共 85 个影响因素。首发及复发 IS 的患者、合并颅内出血的患者，作为混杂因素进行控制。

1.3　研究方法

1.3.1　主成分分析　主成分分析是基于各个变量间的关联关系，利用数据降维的方式，将多个相关联的变量降维成几个少数变量，原始指标的大部分信息均能通过主成分反映出来[15]。本研究中利用主成分分析法进行降维处理，以提高模型的预测效率。

1.3.2　Wide&Deep 模型构建与评价　Wide&Deep 模型中，Wide 模型为以 $y=w^{T}x+b$ 为形式的广义线性模型，Deep 模型为前馈式神经网络模型。结合两类模型的优势，让模型兼具“泛化能力”和“记忆能力”。Wide&Deep 模型结构见图 1。同期构建 SVM、Logistic 回归模型、C5.0 决策树、CART、DNN 等模型，计算模型准确度、灵敏度、召回率等指标对模型进行评价比较，其中准确度是分类正确的样本占总样本个数的占比，精确度是预测为阳性的样本中实际也为阳性的样本占比，灵敏度是实际为阳性的样本中预测也为阳性的样本占比，由于同召回度的计算方式及结果相同，在后续结果展示中省略该指标，综合评价各项模型评价指标选择最佳算法。

将全部原始数据进行分割处理，随机数为 7 和 11，随机分为训练集（1 113 例）、验证集（371 例）、测试集（371 例），其中训练集用来构建和训练模型以发现规律，验证集用来调整模型参数，测试集用来评价最终模型的泛化能力。提取本院 IS 患者的临床信息，以溶栓治疗后患者神经功能是否改善作为输出变量（y），输入变量为用药种类（尿激酶 / 阿替普酶 / 无）、单次剂量、病因分型、溶栓时间窗等 26 个影响因素。

设定模型训练轮次为 100 次，建模完成后，披露建模及模型准确度情况。报告网络层内核心部分 kernel 和偏倚度 bais 收敛性能，描述三维坐标系中数据点的分布情况，判断数据空间分布的聚类性。计算模型准确度、召回度、F-Score 等评价指标，全面评价模型。

1.3.3　寻找 Wide&Deep 模型最优参数　使用

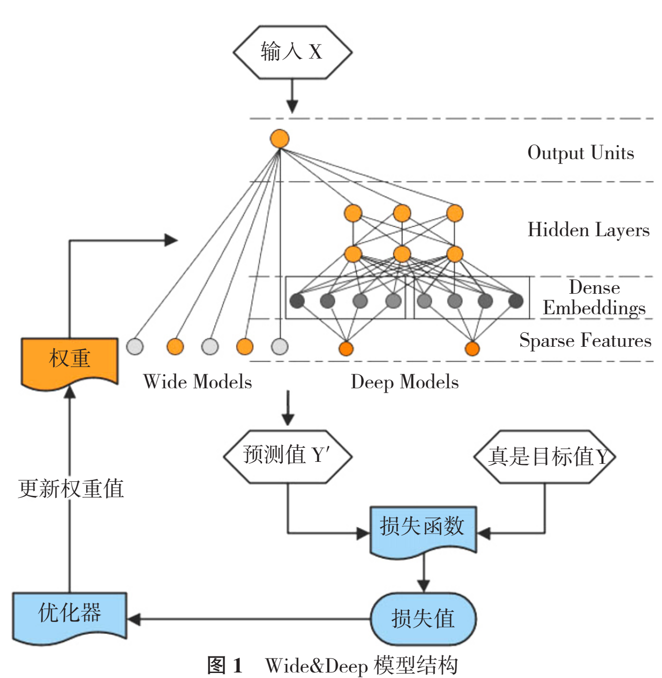


图 1　Wide&Deep 模型结构

Figure 1　The spectrum of the Wide&Deep model

GridSearchCV 方法寻找模型最优参数，在最优参数范围内细微调整 Wide&Deep 模型中激活函数种类，分别为 Relu 和 Sigmoid；隐藏层层数分别为 1 层、3 层、5 层、7 层；隐藏层神经元个数分别为 10、15、30、300。分别设定比较不同条件下模型训练集、验证集、测试集准确度，最终选定网络结构中各个参数。

1.3.4　模型自变量简化　面对 IS 患者具有溶栓的紧迫性、时间的复杂性、用药方案多样性、干扰因素数量多等特点，本研究针对单因素筛选中有统计学意义的因素进行特征工程处理，提取主要影响因素，根据权重大小比较各个因素的重要性程度，以增加模型反应性和实用性。按照各因素权重大小由高到低排序，提取累计百分比 <90% 的变量，构建简化模型并评估其模型准确度。

1.3.5　Wide&Deep 模型外部验证　采用大连医科大学附属第二医院 IS 患者数据建模后，以大连市中心医院的医渡云科研大数据服务器系统中 IS 专病数据库的 IS 患者的临床信息进行外部验证，共提取到 3 925 例患者的临床信息，评估构建的 Wide&Deep 模型的预测及泛化能力，评价指标为模型准确度情况，研究技术路线见图 2。

1.4　统计学方法　采用 SPSS 13.0 统计软件进行数据插补及单因素筛选，针对血小板计数、低密度脂蛋白胆固醇（LDL-C）、同型半胱氨酸等字段构建 CART 树模型填补缺失值。对 85 个影响因素进行单因素筛选，除纳入统计学结果有意义的指标外，还结合神经内科高级职称医师推荐指标，共同作为分析变量纳入主成分分析中，其中符合正态分布的计量资料以（$\bar{x} \pm s$）表示，两组间比较采用两独立样本 t 检验，连续型变量包括年龄、单

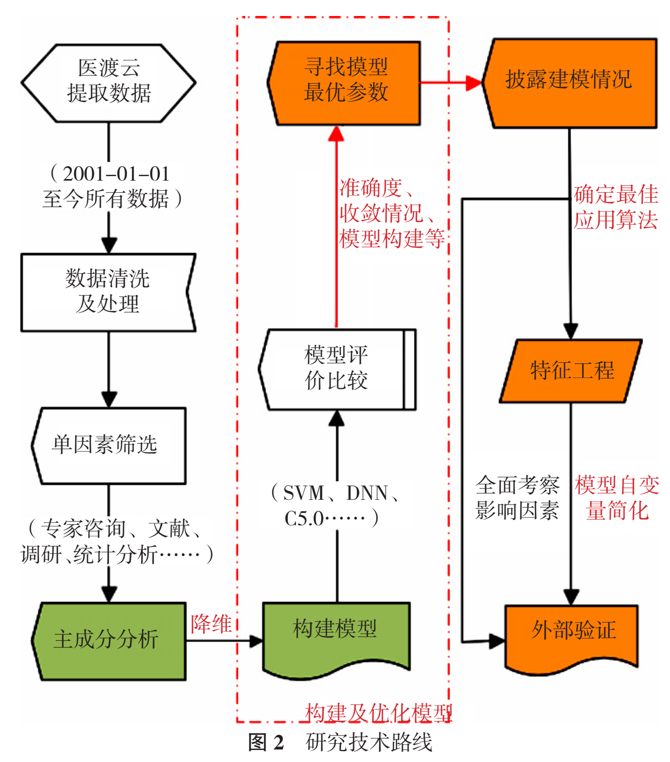


图 2　研究技术路线

Figure 2　Technical route of the study

次剂量、白细胞计数等 54 个指标；计数资料以相对数表示，组间比较采用 χ^2 检验，分类变量包括性别、高血压、是否使用抗血小板药等 31 个指标，以 P<0.05 为差异有统计学意义。利用 Python3.7 进行主成分分析，利用主成分因子构建人工智能模型。使用 Python3.7 中的 Tensorflow 和 Keras 模块构建 DNN 和 Wide&Deep 人工智能模型，寻找模型的最优参数，绘制受试者工作特征（ROC）曲线。同期构建 Logistic 回归模型、SVM、CART 等模型，分别计算 6 种模型的准确度、精确度、召回率等指标进行多种模型的评价。

2　结果

2.1　建模数据集和单因素筛选　纳入的 1 855 例 IS 患者的筛选流程见图 3，研究因素共 85 个。将全部影响因素分为一般特征、用药指标、治疗方式、检查指标、检验指标 5 个方面，结合临床专家意见，纳入 26 个因素构建模型，具体为：年龄、糖尿病患病情况、心房颤动（简称房颤）患病情况、动脉粥样硬化患病情况、出血性疾病史、脑血管病史、高血压患病情况、溶栓药物的单次剂量、用药种类、给药方式、溶栓时间窗、是否使用抗血小板药物、是否使用抗凝药物、是否使用其他活血化瘀药物、是否进行取栓手术、是否桥接治疗、入院 NIHSS 评分、血小板计数、LDL-C、胱抑素、同型半胱氨酸、责任血管狭窄程度、是否有局灶神经功能缺损、梗死部位、梗死面积、病因分型。责任血管狭窄程度：轻度狭窄为狭窄程度 <50%；中度狭窄为狭窄程度 50%~69%，重度狭窄为狭窄程度 70%~99%；局灶神经

功能缺损为IS患者出现无力、麻木、语言障碍的症状[8]。

对照组和神经功能改善组患者年龄、糖尿病患病情况、心房颤动（简称房颤）患病情况、动脉粥样硬化患病情况、出血性疾病史、脑血管病史、高血压患病情况、溶栓药物的单次剂量、用药种类、给药方式、溶栓时间窗、是否使用抗血小板药物、是否使用抗凝药物、是否使用其他活血化瘀药物、是否进行取栓手术、入院NIHSS评分、血小板计数、责任血管狭窄程度、是否有局灶神经功能缺损、梗死部位比较，差异有统计学意义（P<0.05）；对照组和神经功能改善组患者桥接治疗情况、LDL-C、胱抑素、同型半胱氨酸、梗死面积、病因分型比较，差异无统计学意义（P>0.05），见表1。

2.2 主成分分析结果　将26个影响因素降维成2个主成分，方差贡献率分别为65.6%和27.5%，累计方差贡献率为93.1%，见表2。

2.3 不同数据挖掘模型分析比较　采用Logistic回归模型、C5.0树模型、CART树模型、SVM、DNN、Wide&Deep模型评价26个影响因素对IS患者溶栓治疗后神经功能改善的影响，结果显示，Wide&Deep模型准确度、精确度、特异度最高（表3），表示Wide&Deep模型评价最佳。

2.4 Wide&Deep模型参数及准确度情况　分别以Relu、Sigmoid作为激活函数探索模型最佳隐藏层层数，调整隐藏层层数为1、3、5、7层，调整每层神经元节点数为10、15、30、300个，以大连医科大学附属第二医院确诊为IS的患者作为研究对象构建Wide&Deep模型，比较训练集、验证集、测试集模型准确度，最终最优参数为以Sigmoid作为激活函数，隐藏层层数为7层，隐藏层中神经元节点数最佳为15个，见图4。

Wide&Deep模型训练集准确度达到0.816，验证集

表1 对照组和神经功能改善组患者一般特征、用药指标、治疗方式、检查指标、检验指标比较

Table 1 Comparison of general characteristics，medication indicators，treatment methods，examination indicators and test indicators between the control group and the neurological function improvement group

影响因素	对照组（n=619）	神经功能改善组（n=1 236）	检验统计量值	P值
一般特征				
年龄（$\bar{x} \pm s$，岁）	70.8±12.2	68.0±12.6	4.496[a]	<0.001
糖尿病〔n（%）〕	281（45.4）	361（29.2）	47.762	<0.001
冠心病〔n（%）〕	107（17.3）	124（10.0）	19.905	<0.001
房颤〔n（%）〕	126（20.4）	197（15.9）	5.595	0.018
动脉粥样硬化〔n（%）〕	455（73.5）	593（48.0）	109.361	<0.001
出血性疾病史〔n（%）〕	120（19.4）	376（30.4）	25.637	<0.001
脑血管病史〔n（%）〕	457（73.8）	180（14.6）	645.882	<0.001
高血压〔n（%）〕	472（76.3）	810（65.5）	22.194	<0.001
用药指标				
单次剂量（$\bar{x} \pm s$，mg/万U）	17.75±29.29	36.75±33.69	-12.518[a]	<0.001
用药种类〔n（%）〕			681.264	<0.001
无	355（57.4）	53（4.3）		
阿替普酶	167（27.0）	646（52.3）		
尿激酶	97（15.7）	537（43.4）		
给药方式〔n（%）〕			679.514	<0.001
无	355（57.4）	53（4.3）		
静脉滴注	189（30.5）	915（74.0）		
泵注	75（12.1）	268（21.7）		
溶栓时间窗〔M（QR），h〕	1.31（0.95）	1.20（1.65）	-17.519[b]	<0.001
使用抗血小板药物〔n（%）〕	98（15.8）	76（6.1）	60.060	<0.001
使用抗凝药物〔n（%）〕	33（5.3）	38（3.1）	57.858	<0.001
使用其他活血化瘀药物〔n（%）〕	28（4.5）	24（1.9）	56.699	<0.001
治疗方式				
取栓手术〔n（%）〕	149（24.1）	481（38.9）	40.525	<0.001
桥接治疗〔n（%）〕	11（1.8）	19（1.5）	0.149	0.699

影响因素	对照组（n=619）	神经功能改善组（n=1 236）	检验统计量值	P值
检查指标				
入院NIHSS评分〔M（QR），分〕	3（6）	6（10）	-9.293[b]	<0.001
血小板计数（$\bar{x} \pm s$，$\times 10^9$/L）	217.3±58.3	211.7±47.9	2.081[a]	0.038
LDL-C（$\bar{x} \pm s$，mmol/L）	2.44±0.86	2.39±0.73	1.345[a]	0.179
胱抑素（$\bar{x} \pm s$，mmol/L）	1.22±0.61	1.19±0.79	0.789[a]	0.430
同型半胱氨酸（$\bar{x} \pm s$，μmol/L）	14.63±8.10	14.69±7.82	-0.151[a]	0.880
检验指标				
责任血管狭窄程度〔n（%）〕			21.782	<0.001
轻度	18（2.9）	5（0.4）		
中度	1（0.2）	2（0.2）		
重度	8（1.3）	8（0.6）		
未知	592（95.6）	1 221（98.8）		
局灶神经功能缺损〔n（%）〕	288（46.5）	365（29.5）	81.239	<0.001
梗死部位〔n（%）〕			65.713	<0.001
大脑	368（59.5）	563（45.6）		
间脑	4（0.6）	13（1.1）		
小脑和脑干	52（8.4）	87（7.0）		
梗死伴出血	63（10.2）	259（21.0）		
梗死伴脑疝	6（1.0）	15（1.2）		
多发	62（10.0）	82（6.6）		
未知	64（10.3）	217（17.6）		
梗死面积〔n（%）〕			1.386	0.239
非大面积	605（97.7）	1 196（96.8）		
大面积	14（2.3）	40（3.2）		
病因分型			5.356	0.069
心源性	144（23.3）	242（19.6）		
大动脉性	38（6.1）	60（4.9）		
其他	437（70.6）	934（75.6）		

注：NIHSS=美国国立卫生研究院卒中量表，LDL-C=低密度脂蛋白胆固醇；[a]表示t值，[b]表示Z值，余检验统计量值表示χ^2值

表 2 26 个影响因素总方差解释

Table 2 The percentage of total variance explained by 26 influencing factors of thrombolytic effect

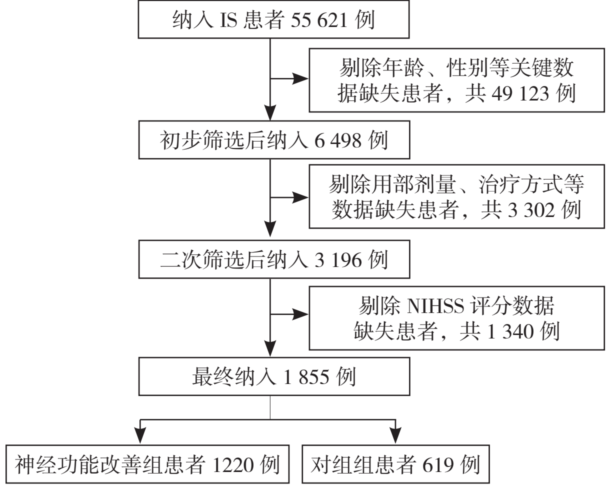
主成分	方差贡献率	累计方差贡献率
1	65.60%	65.60%
2	27.50%	93.10%

表 3 Logistic 回归模型、C5.0 树模型、CART 树模型、SVM 模型、DNN 模型、Wide&Deep 模型评价 IS 患者溶栓治疗后神经功能改善的价值

Table 3 The value of Logistic regression model, C5.0 decision tree arithmetic model, CART tree model, SVM model, DNN model and Wide&Deep model in evaluating the improvement of neurological function in ischemic stroke patients after thrombolytic therapy

数据挖掘模型	准确度	精确度	召回度	特异度	F 指数
Logistic 回归模型	0.693	0.697	0.986	0.142	0.816
C5.0 决策树	0.802	0.814	0.911	0.583	0.860
CART 树模型	0.809	0.813	0.927	0.574	0.866
SVM	0.742	0.754	0.936	0.389	0.835
DNN	0.765	0.854	0.803	0.678	0.828
Wide&Deep 模型	0.815	0.950	0.805	0.850	0.871

注：SVM= 支持向量机，DNN= 深度神经网络



注：IS= 缺血性脑卒中，NIHSS= 美国国立卫生研究院卒中量表

图 3 研究对象纳入流程图

Figure 3 Flow chart of inclusion and exclusion of the participants

准确度达到 0.828，测试集准确度达到 0.844，模型准确度较高。模型准确度曲线如图 5。

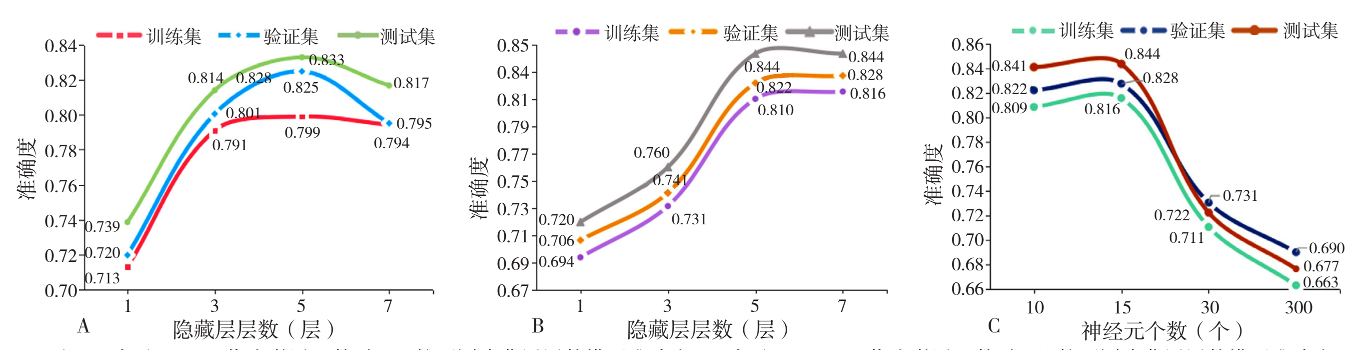
大连医科大学附属第二医院 IS 患者中训练组 ROC 曲线下面积为 0.753，测试组曲线下面积为 0.793（图 6），表明 Wide&Deep 模型具有良好的预测性能和泛化能力，模型没有出现过拟合情况。

2.5 经特征工程的自变量简化及外部验证结果 IS 患者溶栓治疗后神经功能改善影响因素的特征工程分析结果显示，用药种类、给药方式和用药剂量的重要性排序均在前列，重要性排序由大到小分别为：脑血管病史、用药种类、给药方式、单次剂量、动脉粥样硬化、溶栓时间窗、是否使用抗凝药物和活血化瘀药物等，表明药物因素在 IS 患者治疗效果中占重要作用（表 4）。为更好的应用模型，进行模型自变量简化，得出 Wide&Deep 模型准确度为 0.819，外部验证模型准确度为 0.800，表示 Wide&Deep 模型后仍具有较好的泛化性及预测性能。

3 讨论

3.1 人工智能算法在药物选择中的应用 本研究采用 Wide&Deep 模型能够解决从患者溶栓影响因素到临床中患者神经功能改善的有效预测，模型测试集预测准确率达到 0.844。IS 患者溶栓影响因素中包含用药种类、单次剂量、给药方式等治疗方案，通过结合患者个体特征和用药方案指标，就可以把不同溶栓用药方案的治疗效果推荐给临床医生，供临床决策辅助使用，让更多患者获益。本研究构建出 6 种模型（表 2），综合模型评价指标，选择 Wide&Deep 神经网络为模型算法，作为本研究的人工智能模型。本研究引入了 Wide&Deep 神经网络算法，在既往研究中尚少见[16]，其模型核心是将深度神经网络具备的泛化性和线性模型的记忆性能相结合，较大程度上提升模型的预测性能。

3.2 IS 溶栓治疗结果的影响因素分析 目前研究多集中在溶栓药物的临床效果观察，溶栓药物的效果和安全性分析，溶栓效果影响因素研究等方面，样本量多集中在数十到数百不等。临床疗效观察中，多采用前瞻性研



注：A 表示以 Relu 作为激活函数时，比较不同隐藏层层数模型准确度；B 表示以 Sigmoid 作为激活函数时，比较不同隐藏层层数模型准确度；C 表示为每层神经网络不同神经元个数模型的准确度

图 4 优化 Wide&Deep 模型的最优参数

Figure 4 Optimal parameters of the Wide&Deep model after optimization

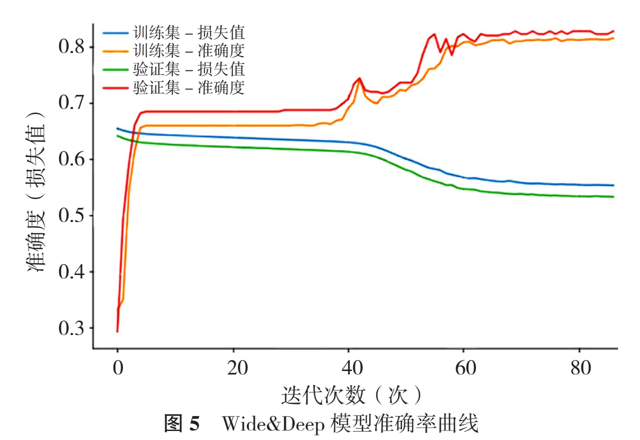


图5　Wide&Deep 模型准确率曲线

Figure 5　Accuracy curve of the Wide&Deep model

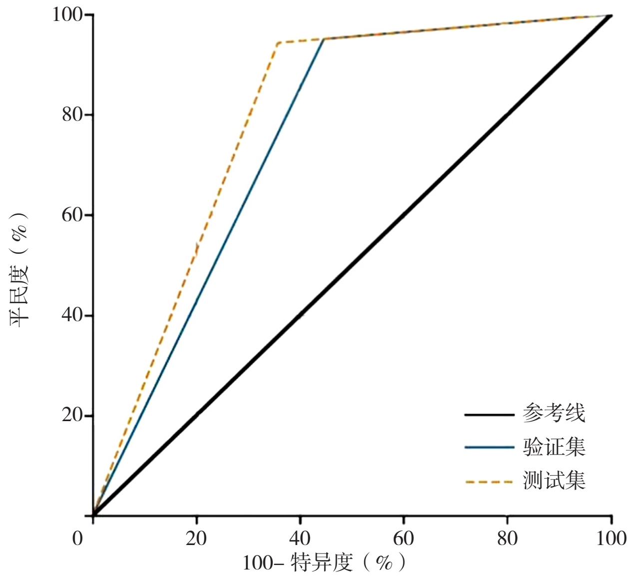


图6　Wide&Deep 模型预测训练组和测试组 IS 患者溶栓治疗后神经功能改善的 ROC 曲线

Figure 6　ROC curves of the Wide&Deep model in predicting the neurological function improvement in ischemic stroke patients in training group and test group after thrombolytic therapy

究，分为基础治疗组和加用溶栓药物组，针对疗效结果进行统计比较，得出试验结论，结果多是给出溶栓药物的疗效，针对溶栓治疗方法的选择，包括给药剂量和方式等探讨尚未见报道。在影响因素探索方面，现有研究多采用多因素 Logistic 回归分析方法[10-11]，通过单因素筛选，得到有意义的影响因素后进行多因素回归，得到最终的影响因素，但基于大数据算法和真实世界全数据研究的相关文献尚少见。本研究通过预测模型程序，输入患者的相关因素，其中包括穷举溶栓药物治疗方案，包括溶栓药种类、单次剂量、给药方式等，就能通过模型寻找出最佳给药方案组合，为临床医生的决策提供理论参考，提升缺血性脑卒中治疗效果，实现对缺血性脑卒中患者的个体化精准治疗，对减轻疾病社会负担具有积极意义。

表4　经特征工程得到影响因素重要性排序结果

Table 4　The importance ranking of influencing factors for thrombolytic effect obtained through feature engineering

影响因素	权重	权重占比（%）	累计权重占比（%）
是否有脑血管病史[a]	988.87	33.46	33.46
用药种类[a]	670.44	22.68	56.14
给药方式[a]	510.73	17.28	73.42
单次剂量[a]	142.85	4.83	78.25
动脉粥样硬化[a]	116.09	3.93	82.18
溶栓时间窗[a]	96.16	3.25	85.43
是否使用抗凝药物[a]	58.94	1.99	87.43
是否使用其他活血化瘀药物[a]	54.35	1.84	89.27
糖尿病	48.97	1.66	90.92
入院 NIHSS 评分	44.86	1.52	92.44
是否使用抗血小板药物	41.52	1.40	93.85
是否进行取栓手术	41.39	1.40	95.25
是否患出血性疾病	25.97	0.88	96.13
梗死部位	25.68	0.87	96.99
高血压	22.44	0.76	97.75
责任血管狭窄	22.31	0.75	98.51
年龄	20.21	0.68	99.19
房颤	5.61	0.19	99.38
血小板计数	4.93	0.17	99.55
是否神经功能缺损	4.72	0.16	99.71
病因分型	4.65	0.16	99.87
低密度脂蛋白胆固醇	1.81	0.06	99.93
梗死面积	1.39	0.05	99.97
胱抑素	0.62	0.02	99.99
是否桥接治疗	0.15	0.01	100.00
同型半胱氨酸	0.02	0	100.00

注：[a] 表示累计权重占比 <90% 的因素，为重要性较强影响因素

本研究经特征工程模型自变量简化后，得到 IS 患者治疗后影响其溶栓效果的重要因素，分别为用药种类、给药方式、单次剂量及是否有脑血管病史等。其中，是否有脑血管病史在特征工程结果中得到权重为 988.87，居于第一位，有研究表明，有心脑血管疾病史的患者更易发作 IS，如出血性脑血管病、缺血性脑血管病、脑供血不足、高血压脑病、动脉粥样硬化或狭窄等，易引起脑血管血流变化，是 IS 发作的重要诱因[17]。在分析影响 IS 患者溶栓治疗效果因素的单因素筛选时，得到用药因素中用药种类、给药方式、单次剂量在神经功能改善组和对照组中差异明显（$P<0.001$），表明用药因素对于患者溶栓治疗效果的影响均较大。由特征工程模型自变量简化结果得到，用药种类权重居于第二位，权重大小为 670.44，所占百分比为 22.68%。本研究中针对静脉溶栓药物，研究团队纳入阿替普酶和尿激酶作为主要研究药物，对 IS 发作在 4.5 h 以内的患者，应按照适

应证、禁忌证等严格筛选患者，并给予重组人组织型纤溶酶原激活物（rt-PA）；IS 发作 6 h 内给予尿激酶，相对安全有效，但其适应证、禁忌证等均有待更新与修订[14]。本研究能够根据患者个体化特点，针对不同溶栓时间窗、患者 NIHSS 评分、血糖水平等给出临床用药种类的推荐方案。给药方式权重居于第 3 位，权重大小为 510.73，所占百分比为 17.28%，在给予患者静脉溶栓过程中，给药方式分为两种：静脉推注或使用泵入，使用泵入时要求抽取 10% 药物在 1 min 内静脉推注，其余药物在 1 h 内静脉泵入完成，因此给药方式也是研究团队想要给出的药物溶栓治疗方案的因素之一，提供用药方式参考，以达到较好溶栓效果。单次剂量权重居于第 4 位，权重大小为 142.85，所占百分比为 4.83%，诊治指南中推荐 rt-PA 给药正常剂量为 0.9 mg/kg，或给予小剂量 rt-PA 0.6 mg/kg，小剂量 rt-PA 出血风[14]险小于正常剂量，但未见降低患者治疗后致残率，需考虑患者病情严重情况、出血情况等进行个体化决策，使用尿激酶作为静脉溶栓药物时，推荐使用应用剂量为 100~150 U，本研究构建人工智能模型，旨在纳入新的患者时能够推荐患者最佳溶栓药物剂量。当有新患者进入人工智能模型后，模型经算法运行，可结合患者自身情况给出治疗时用药种类（阿替普酶和尿激酶）、单次剂量及给药方式(使用泵或静脉滴注)等，辅助临床决策。

3.3　人工智能与未来医药　伴随大数据与人工智能产业不断升温，人工智能技术正逐渐渗透进人们的日常生活中。在医疗卫生领域，人工智能正进一步拓宽其发展空间及应用前景，展现出愈发重要的应用价值，助推医疗事业正向发展，如病理诊断、辅助诊疗、医学影像识别、药物研发等[18]。相信在医疗卫生事业发展中，人工智能将与医疗进一步融合，成为辅助医护人员诊疗决策的重要支撑[19]。

3.4　IS 治疗紧迫性及社会负担　卒中对全球经济影响巨大，根据《心脏疾病和卒中统计 -2020 更新》统计，由脑卒中带来直接和间接经济损失达 4 550 万美元，预估直接医疗费用高达 2 800 万美元[20]。在我国，脑卒中是仅次于癌症的第二位致死性疾病。本研究针对 IS 患者的不同特征，为其选择溶栓方案提供参考，深入研究影响溶栓的因素，为IS卫生经济学决策提供数据支持，并为不同特征患者优选出临床决策方案，从而以较低成本获得较高的治疗收益。

本研究尚存在一定的不足，本研究纳入了 1 855 例 IS 患者的临床数据构建模型，Wide&Deep 模型的准确度可达到 81.5%，模型准确度及预测性能较好，若纳入更多患者数据，模型准确度可以进一步提高。此后，本团队将纳入更多中心数据进行外部验证，进一步对模型的预测性能进行评估，对模型持续调整。

综上所述，人工智能算法结果显示，影响溶栓效果的因素排序由大到小分别为：脑血管病史、用药种类、给药方式、单次剂量、动脉粥样硬化、溶栓时间窗、是否使用抗凝药物和活血化瘀药物等，能够为临床决策提供参考实现对 IS 患者的个体化精准治疗，对减轻疾病社会负担具有积极意义。

作者贡献：沈惠文负责数据收集和整理，统计学处理，并撰写论文初稿；林永忠提出主要研究指标；陈淑良负责纳排标准的制定张立红负责究对象的选取；马春野负责论文修订；马得原负责数据收集和整理；张策负责研究的质量控制及审校，并对研究负责；所有作者确认了论文的最终稿。

本文无利益冲突。

参考文献

［1］CHENG H T，KOC L，HARMSEN J，et al. Wide&deep learning for recommender systems［EB/OL］.［2021-04-02］. https://arxiv.org/abs/1606.07792.

［2］李小敏，曲扬，张少霆，等 . 人工智能技术在骨肌系统影像学方面的应用［J］. 上海交通大学学报：医学版，2021，41（2）：262-266. DOI：10.3969/j.issn.1674-8115.2021.02.022.
LI X M，QU Y，ZHANG S T，et al. Overview of the application of artificial intelligence to radiology of the musculoskeletal system［J］. Journal of Shanghai Jiao Tong University：Medical Science，2021，41（2）：262-266. DOI：10.3969/j.issn.1674-8115.2021.02.022.

［3］施昭，张龙江 . 人工智能在脑血管病影像学中的应用［J］. 中华放射学杂志，2021，55（8）：793-798. DOI：10.3760/cma.j.cn112149-20210616-00564.
SHI Z，ZHANG L J. Artificial intelligence in cerebrovascular disease imaging［J］. Chinese Journal of Radiology，2021，55（8）：793-798. DOI：10.3760/cma.j.cn112149-20210616-00564.

［4］吴韵阳，高键东，吴及 . 人工智能影像技术在卒中诊疗中的应用［J］. 中国卒中杂志，2021，16（7）：651-656. DOI：10.3969/j.issn.1673-5765.2021.07.003.
WU Y Y，GAO J D，WU J. Application of artificial intelligence imaging in diagnosis and treatment of stroke［J］. Chinese Journal of Stroke，2021，16（7）：651-656. DOI：10.3969/j.issn.1673-5765.2021.07.003.

［5］李小东，尹晶，于昌玉，等 . Fogarty 球囊导管联合溶栓药物治疗断肢再植术后联合血栓形成的临床效果［J］. 广西医学，2017，39（8）：1246-1248. DOI：10.11675/j.issn.0253-4304.2017.08.40.

［6］黄义松 . 尿激酶溶栓治疗急性脑梗死的临床效果观察［J］. 医药前沿，2019，9（13）：58-59.
HUANG Y S. Clinical observation of urokinase thrombolysis in the treatment of acute cerebral infarction［J］. Journal of Frontiers of Medicine，2019，9（13）：58-59.

［7］鲁庆波，王雪梅，汪鸿浩，等 . MR 筛选下 4.5~9 h 缺血性脑卒中阿替普酶静脉溶栓联合标准Ⅱ级预防治疗的效果及安全性分析［J］. 中国医学装备，2021，18（6）：64-67. DOI：10.3969/J.ISSN.1672-8270.2021.06.017.
LU Q B，WANG X M，WANG H H，et al. Analysis on the efficacy

and safety of intravenous thrombolytic therapy with Alteplase combined with standard grade Ⅱ preventive treatment in 4.5-9 hours IS under MR screening[J]. China Medical Equipment, 2021, 18(6): 64-67. DOI：10.3969/J.ISSN.1672-8270.2021.06.017.
[8] 王正中，王丽，冉立群，等. 重组人尿激酶原在STEMI溶栓治疗中的安全性和疗效分析[J]. 中国生化药物杂志，2016，36(6): 170-172.
[9] 刘啸，张磊，陶伟. 急性脑梗死患者rt-PA静脉溶栓治疗短期预后的影响因素及血清Cys C水平的预测价值[J]. 山东医药，2021，61(25): 49-51. DOI: 10.3969/j.issn.1002-266X.2021.25.012.
[10] 孙德金，许铁. 急性脑梗死溶栓后脑出血的危险因素Logistic回归分析[J]. 西南国防医药，2017，27（12）：1268-1271. DOI：10.3969/j.issn.1004-0188.2017.12.006.
[11] 张莉宏，田建荣，负红雨，等. Logistic回归模型回顾性分析脑卒中溶栓后的死亡影响因素[J]. 糖尿病天地，2021，18(6): 134-135. DOI：10.12083/SYSJ.2018.06.173.
[12] 王馨，高嵩芹，于龙娟，等. "一站式"护理救治模式提升缺血性脑卒中患者救护质量的效果[J]. 解放军护理杂志，2015，32（20）：57-59，63. DOI：10.3969/j.issn.1008-9993.2015.20.015.
[13] HACKE W，KASTE M，BLUHMKI E，et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke[J]. N Engl J Med，2008，359（13）:1317-1329. DOI：10.1056/NEJMoa0804656.
[14] 中华医学会神经病学分会，中华医学会神经病学分会脑血管病学组，彭斌，等. 中国急性缺血性脑卒中诊治指南2018[J]. 中华神经科杂志，2018，51（9）：666-682.
[15] 张娜，张建丰，李涛. 西安地区冬小麦灌溉水分生产率影响因素的主成分分析[J]. 西安理工大学学报，2021，37(1): 9-15. DOI：10.19322/j.cnki.issn.1006-4710.2021.01.002.
[16] BONKHOFF A K，GREFKES C. Precision medicine in stroke: towards personalized outcome predictions using artificial intelligence[J]. Brain，2022，145（2）：457-475. DOI：10.1093/brain/awab439.
[17] 刘宁，张高才，徐建可，等. 椎基底动脉延长扩张症与急性缺血性脑卒中患者预后的相关性[J]. 中华老年心脑血管病杂志，2020，22（8）：793-796. DOI：10.3969/j.issn.1009-0126.2020.08.003.
LIU N，ZHANG G C，XU J K，et al. Relationship between vertebrobasilar artery dolichoectasia and outcome in acute ischemic stroke patients[J]. Chinese Journal of Geriatric Heart Brain and Vessel Diseases，2020，22（8）：793-796. DOI：10.3969/j.issn.1009-0126.2020.08.003.
[18] 王笛，赵靖，金明超，等. 人工智能在医疗领域的应用与思考[J]. 中国医院管理，2021，41（6）：71-74.
[19] 徐亮，阮晓雯，李弦，等. 人工智能在疾病预测中的应用[J]. 自然杂志，2018，40（5）：349-354. DOI：10.3969/j.issn.0253-9608.2018.05.004.
[20] 张科宝. 急性缺血性脑卒中出院时神经功能预后模型的构建[D]. 汕头：汕头大学，2021.

（收稿日期：2023-01-07；修回日期：2023-03-01）
（本文编辑：毛亚敏）